NewsWise

Anshul Modh, Irfan Radarma, Li Liu, Xiangling Liu

Content

1. Product Concept
2. Product AI Canvas
3. Product Team
4. Value of Data
5. Data Flywheel
6. Model Selection
7. Model Metrics
8. MVP Development and Lessons Learned
9. Proposed Architecture
10. MVP Demo
11. Future Goals

1. Product Concept/Overview

NewsWise is a web application that helps combat the spread of misinformation by using bias detection and misinformation prediction to evaluate news articles for trustworthiness. It targets media companies, political organizations, researchers, and government agencies and gives them a host of tools to evaluate the credibility of the sources they use.

2. Product AI Canvas

Opportunity

The spread of misinformation and biased news is a growing problem which results in $78 billion of losses annually. There is a need for a tool that can help identify and combat this issue. NewsWise provides an opportunity to address this problem by utilizing powerful language models to analyze text for biases, inconsistencies, and truthfulness so that businesses and individuals can get a second opinion about the content they consume.

Solution

NewsWise is a content analysis tool that automatically generates a summary of the facts and claims in a news article, analyzes text for biases and inconsistencies, and assesses the overall truthfulness and bias of a text. It utilizes advanced language models, including GPT and RoBERTa, to provide accurate and reliable results that are based on the semantic understanding of language.

Consumers

NewsWise targets journalists, researchers, and organizations that need to analyze large amounts of text quickly and where their business depends on the accuracy of the data they use. Specifically, businesses where the impact of proliferating misinformation is high.

Data

A compilation of real/fake news articles from FakeNewsNet and OnionOrNot is used to train and validate fake news prediction. An expert-annotated dataset of sentences and their bias for bias detection is used. Claim summarization uses a text completion model (GPT) trained on verified texts.

Strategy

We will offer a superior product that is unparalleled in the features and value it offers. Similar to grammarly, News-wise will position itself as a bias and misinformation mitigation tool for businesses. By collecting user feedback, News-wise will continue to iterate on its modeling strategy.

Policy & process

Since we utilize generative models, we are prone to unanticipated model behavior. Rigorous runtime validation needs to be implemented to detect and mitigate when this happens. Transparent documentation to users should be a priority.

Transfer learning

NewsWise improves language model accuracy and efficiency by fine-tuning on relevant datasets and user feedback. We will experiment with a mix of heavily fine-tuned models and few-shot learning techniques to generalize to a breadth of topics.

Success criteria

NewsWise's success can be measured by model accuracy and reliability and its ability to reason about bias in text. This will be tracked with user feedback and validation performance of the models on topics it has not seen before such as the OnionOrNot dataset.

3. Product Team

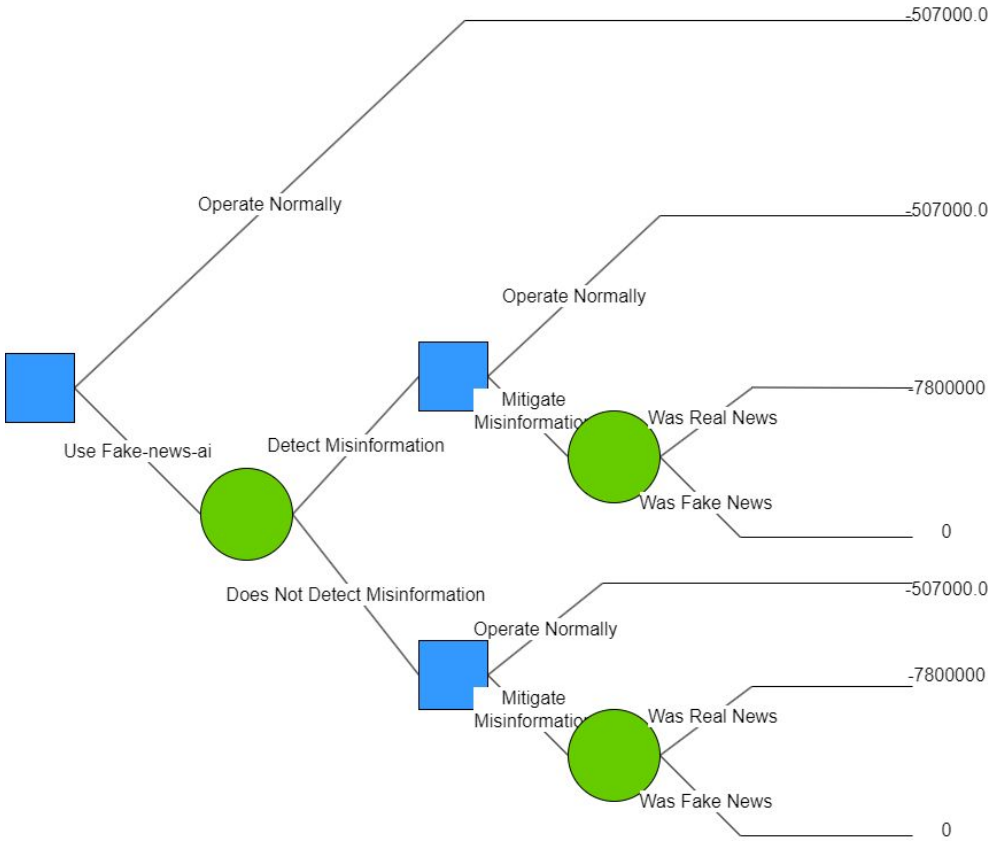
Product Manager	Define the product vision, strategy, and roadmap. Conduct market research, identify customer needs and pain points, and prioritize features and functionalities.
Data Engineer	Develop and maintain NewsWise's data infrastructure, including data sourcing, cleaning, preprocessing, and database design and management.
Data Scientist	Develop ML algorithms for text analysis, train and fine-tune GPT and RoBERTa models for accuracy, reliability, and scalability.
ML Engineer	Implement machine learning models in production, ensuring scalability and reliability.
Back-end Developer	Build and integrate ML models into NewsWise back-end using Python for scalability, security, and reliability. Develop the front-end of the NewsWise platform using React and Material UI.

4. Value of Data - 1

- $78 Billion lost annually due to misinformation (Study by Cybersecurity firm CHEQ and The University of Baltimore)
- Assumptions:
  - We are a large business comprising of 0.1% of the worldwide media market analyzing our content's economic impacts (not directly related to our business model)
  - Cost of spreading misinformation is -10x worse than profit gained from real news
  - Likelihood of fake news is 15%
  - Mitigation when discovering misinformation results in $0 of revenue

4. Value of Data - 2

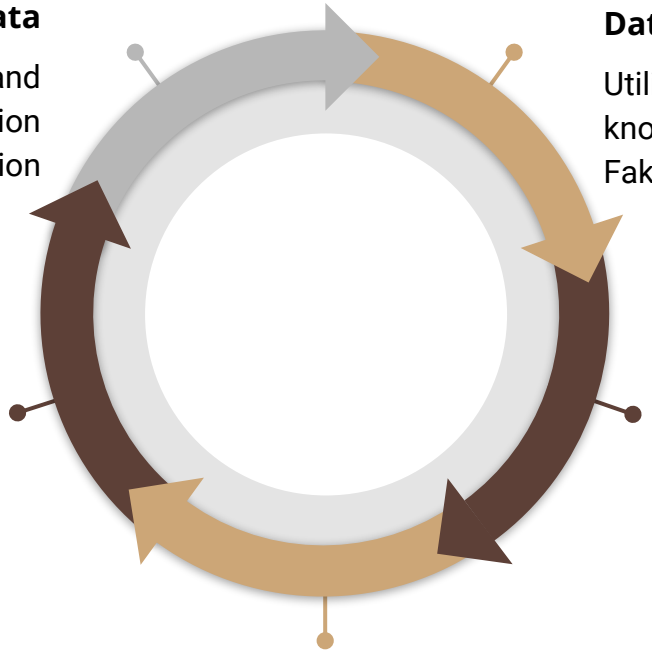
- Data is valuable when sensitivity and specificity are over 0.76
- Current model only achieves this on FakeNewsNet, saves business $45,240 yearly
- On OnionOrNot, model performs with 0.79 sensitivity and 0.54 specificity
    - Could be incorporated into the training set in the future to improve performance



5. Data Flywheel

More Users -> More Data

Gather news articles and feedback on bias detection and misinformation



Data Collection

Utilize user feedback and known data sources such as FakeNewsNet and OnionOrNot

Deploy Model -> Improve User Experience

Deploy improved GPT model to OpenAI and/or improved RoBERTa model to backend microservices for user feedback

Data Preparation + Human Scoring

Curate the data for relational store, truncate lengths, and split sentences; expert annotate portions of the dataset to compare accuracy of user feedback and for training

Model Training

Fine-Tune RoBERTa and GPT 3.5 on downstream tasks, compare performance with prior runs

6. Model Selection - 1

Bias Scoring

- Give an overall score of the amount of bias in a piece of text
- Helps understand the neutrality of writing

Pre-Trained RoBERTa

Fake News Classification

- Predict whether a news article could be fake or not
- Output a confidence score to aid interpretability

Fine-Tuned GPT 3.5

Claim Summarization & Bias Reasoning

List the most important claims in a piece of text and understand if they hinge on any bias

Pre-Trained GPT 3

6. Model Selection - 2

Model	Task	Dataset	Model Output
Pre-Trained RoBERTa	Bias Detection	BABE (Bias Annotations By Experts)	Bias (0 - 1)
Fine-Tuned GPT 3.5	Fake News Classification	Fake News Net	Fake (0,1)
Pre-Trained GPT 3	Text Completion	Books + Wikipedia + Verified Online Sources	Claim Summarization & Bias Reasoning

7. Model Metrics

Bias Detection - Metrics

- Fine-Tuning Dataset: BABE
  - expert-annotated sentences
  - Training Data: 3700 sentences
  - Validation Data: 1700 sentences
- Metric: Macro F1
  - 5-fold cross validation results with standard error shown
- Baseline: XGBoost trained on TF-IDF features
- “Neural Media Bias Detection Using Distant Supervision With BABE - Bias Annotations By Experts”

Model	Macro-F1	Standard Error Across Folds
Baseline	0.511	0.008
RoBERTa	0.798	0.022

Fake News Classification - Metrics

- Fine-Tuning Dataset: FakeNewsNet
  - ~ 61,000 articles
  - Training Data:
    - FNN-Small: 100 samples
    - FNN-Large: 10,000 Samples
  - Validation Data:
    - OnionOrNot: 1000 Samples
- Metric: Accuracy, Macro Precision, Macro Recall

Model	Training Accuracy	Validation Accuracy	Validation Precision	Validation Recall
Fine-Tuned on FNN-Small	95%	69%	67%	66%
Fine-Tuned on FNN-Large	99%	61%	80%	51%

8. Lessons Learned

8. Lessons learned - Model

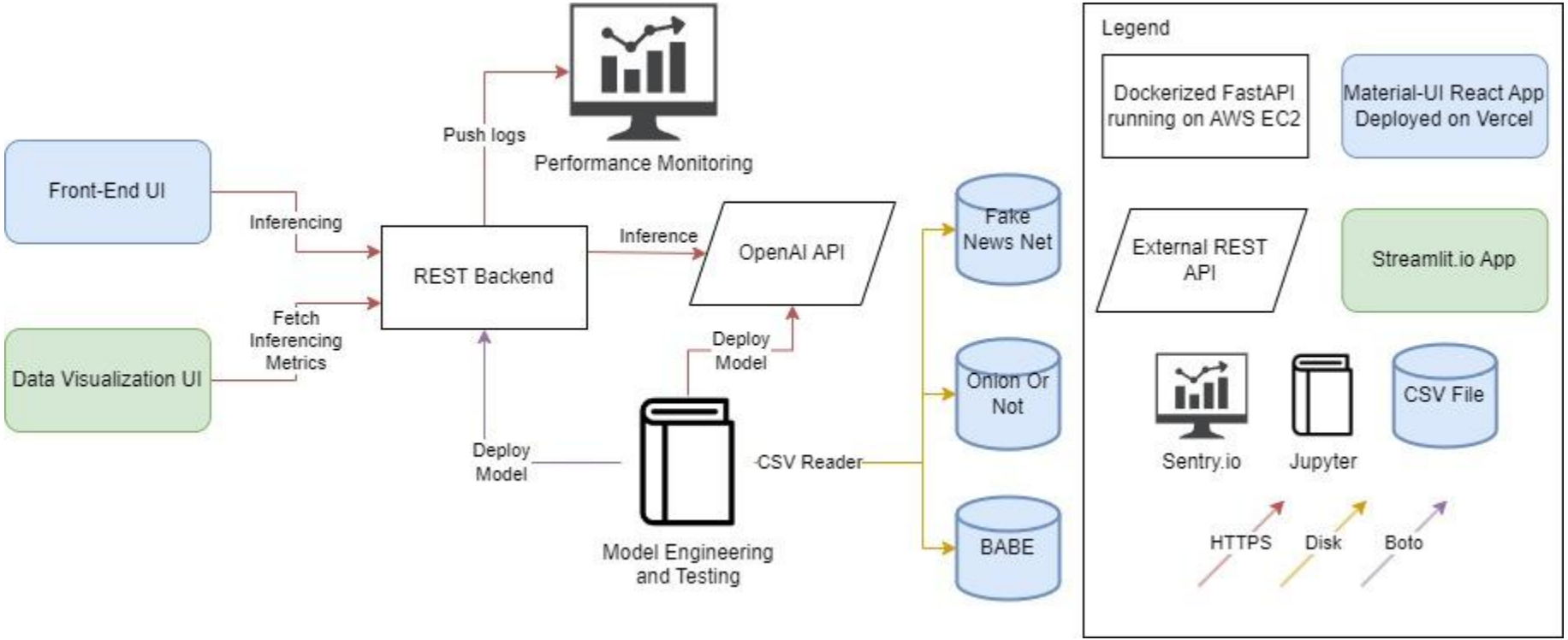
- Fake News Prediction
  - Overfitted to political articles on FakeNewsNet, performs much worse on Onion dataset
  - Few-Shot learning may generalize better
    - F1 Score of FNN-Small fine-tuned model was better than on FNN-Large
    - FNN-Small: 0.66 versus FNN-Large: 0.40
  - Should incorporate samples from OnionOrNot in training data for production models
- Claim Summarization and Bias Reasoning
  - Generative models need robust error handling
    - Unexpected output can cripple functionality, can be rectified with example outputs in the prompt

8. Lessons learned - MVP Development

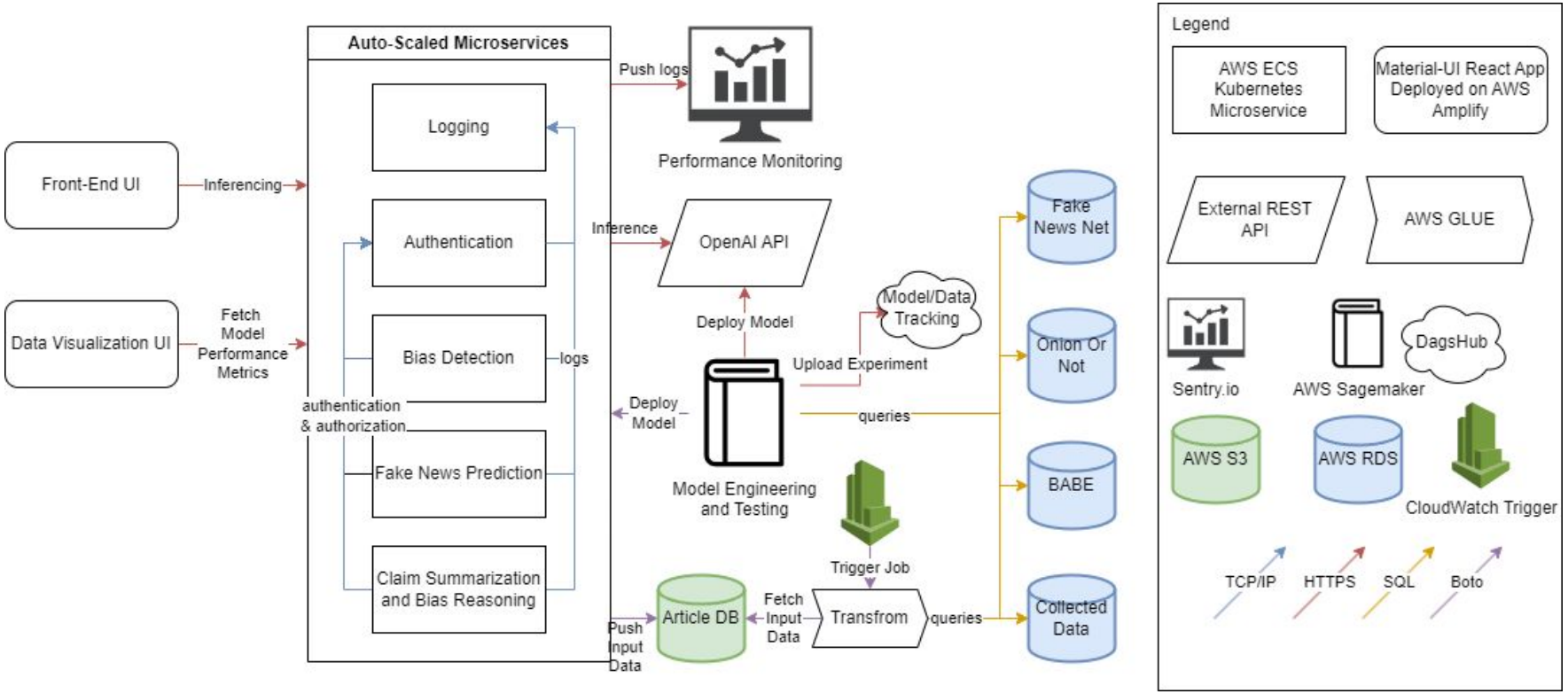
- Backend is tough to scale, currently can only support a single request at a time
  - Microservices architecture could improve scalability
- Processing entire article at once has a latency hit, should experiment with processing pieces at a time (sentence level reasoning rather than entire article)
- Users would get more value out of bias detection on a sentence by sentence basis

9. Proposed Architecture

Current Architecture



Proposed Architecture



Architectural Considerations

Scaling	Horizontally scale with a microservices architecture
Data Collection/Curation	Upload data to a data lake in S3 with AWS Glue job triggers to put it into a structured database
Pipelines	AWS Glue running Pyspark jobs; triggered by AWS Cloudwatch
Labeling	Label dataset based on user feedback on the application
Model Deployment	Bias detection model can be packaged with Onnx; GPT models managed by OpenAI
Monitoring	Sentry.io for software monitoring; log model performance alongside input data and visualize with a tableau-powered dashboard
Testing/Explainability	Compute model confidences, A/B Testing in production
Governance	DVC & DagsHub for data and model; Models can have a high degree of unpredictability on our tasks, need to be transparent with our users on how to best utilize the information

10. MVP Demo

MVP Demo

Product website:

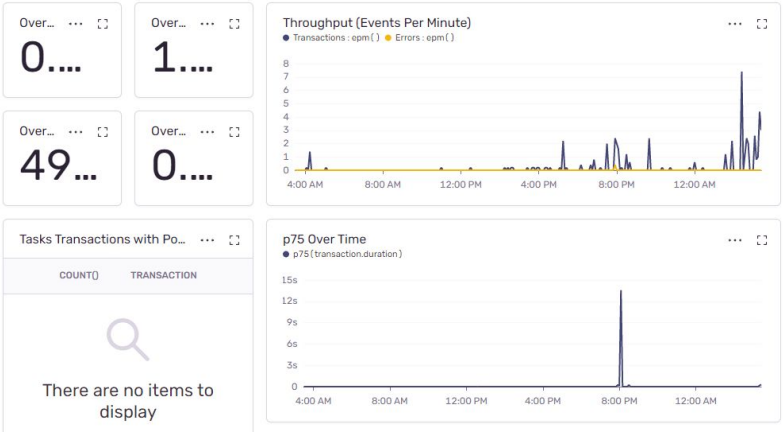
https://www.news-wise-ai.com/

Input data track:

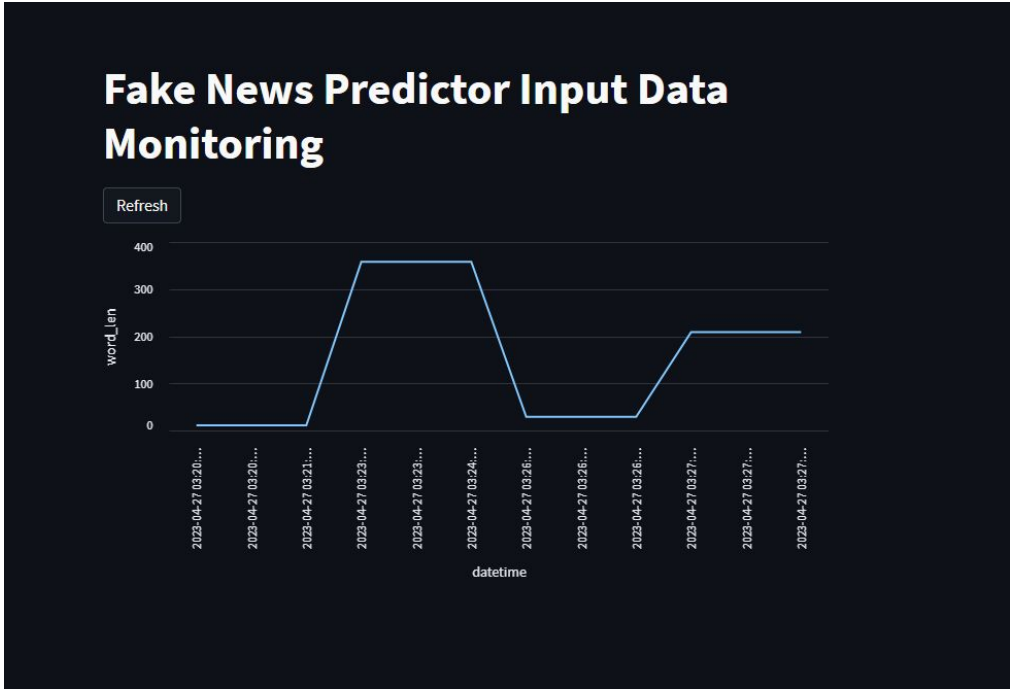
https://newswise-cmu-streamlit-data-monitoring-data-monitoring-agkqns.streamlit.app/

Monitoring

Sentry.io Performance Monitoring



Streamlit.io Input Monitoring



11. Future Goals

- Setup CI/CD for front and back end
- Automatically deploy models with a script
- Highlight sentences with high bias rather than applying a score to the whole text
- Add data storage to collect user input and feedback

References

Timo Spinde, Lada Rudnitckaia, Jelena Mitrovic, Fe-´ lix Hamborg, Michael Granitzer, Bela Gipp, and Karsten Donnay. 2021b. Automated identification of bias inducing words in news articles using linguistic and context-oriented features. Information Processing & Management, 58(3):102505.

Timo Spinde, Manuel Plank, Jan-David Krieger, Terry Ruas, Bela Gipp, and Akiko Aizawa. 2021. Neural Media Bias Detection Using Distant Supervision With BABE - Bias Annotations By Experts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1166–1177, Punta Cana, Dominican Republic. Association for Computational Linguistics.